Structure, Vol. 12, 397–407, March, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2004.02.009

Crystal Structure of RumA, an Iron-Sulfur Cluster Containing *E. coli* Ribosomal RNA 5-Methyluridine Methyltransferase

Tom T. Lee, Sanjay Agarwalla, and Robert M. Stroud*
Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco California 94107

Summary

RumA catalyzes transfer of a methyl group from *S*-adenosylmethionine (SAM) specifically to uridine 1939 of 23S ribosomal RNA in *Escherichia coli* to yield 5-methyluridine. We determined the crystal structure of RumA at 1.95 Å resolution. The protein is organized into three structural domains: The N-terminal domain contains sequence homology to the conserved TRAM motif and displays a five-stranded β barrel architecture characteristic of an oligosaccharide/oligonucleotide binding fold. The central domain contains a [Fe_4S_4] cluster coordinated by four conserved cysteine residues. The C-terminal domain displays the typical SAM-dependent methyltransferase fold. The catalytic nucleophile Cys389 lies in a motif different from that in DNA 5-methylcytosine methyltransferases. The electrostatic potential surface reveals a predominately positively charged area that covers the concave surface of the first two domains and suggests an RNA binding mode. The iron-sulfur cluster may be involved in the correct folding of the protein or may have a role in RNA binding.

Introduction

Posttranscriptional RNA modifications are found in all types of RNA throughout the biosphere. At least 96 different types of nucleotide modifications have been identified (Rozenski et al., 1999). Thirty-five modified nucleotides (11 in 16S, 24 in 23S) have been identified in *E. coli* rRNAs (Del Campo et al., 2001; Limbach et al., 1994). Most of the modifications are located at functionally important sites of the ribosome, such as the peptidyl transferase reaction center, polypeptide exit tunnel, intersubunit bridges, and various tRNA binding sites (Decatur and Fournier, 2002). These posttranscriptional modifications are performed by enzymes that are highly specific for their target substrates. Investigating the structures of these enzymes provides us a venue for understanding not only the chemistry of catalysis but also the specific RNA recognition by proteins.

A common modification found in RNA is the methylation of uridine to 5-methyluridine (m^5U). In *E. coli*, m^5U exists at three sites: tRNA m^5U54, 23S rRNA m^5U747, and 23S rRNA m^5U1939. The enzyme responsible for tRNA m^5U54, RUMT, has been well studied (Kealey et al., 1994). A bioinformatics study identified two open reading frames in the *E. coli* genome, namely ygcA and ybjF, as potential m^5U methyltransferase (MTase) genes (Gustafsson et al., 1996). They were proposed to be responsible for the formation of m^5U747 and m^5U1939 of the 23S rRNA. Subsequent characterization of ygcA gene product confirmed that it specifically methylates U1939. It was named RumA (Agarwalla et al., 2002).

U1939 is close to the peptide exit tunnel of the ribosome and is located in a conserved sequence (1930–1969) in domain IV of 23S rRNA, which includes stem H70, the 1942 loop (containing U1939), stem H71, and the 1953 loop. In the crystal structure of the 70S ribosome, this segment is seen to bridge with H44 of 16S rRNA through minor groove-minor groove interactions (Yusupov et al., 2001). Furthermore, the U1939-containing loop is inserted into the major groove of the acceptor stem of A site tRNA at the CCA tail. This interaction might have an as yet undetermined but important functional role in translation.

S-adenosylmethionine (SAM)-dependent m^5U and 5-methylcytidine (m^5C) MTases, including RumA, share a common catalytic mechanism which utilizes a cysteine nucleophile on the enzyme to form a Michael adduct with the pyrimidine base, thereby activating C5 for accepting a methyl group from SAM (Figure 1) (Kealey et al., 1994). Structures of two DNA m^5C MTases, *M. Hha*I and *M. Hae*III, including their complexes with cofactor and DNA substrates have been reported (O'Gara et al., 1996; Reinisch et al., 1995), and the structure of one RNA m^5C MTase, Fmu, has recently been determined (Foster et al., 2003). However, heretofore no structure for SAM-dependent m^5U MTase has been reported.

The RumA subfamily of the m^5U MTases is distinguished from the RUMT and YbjF subfamilies by the presence of an N-terminal extension which has sequence homology to the "TRAM" domain. TRAM is a domain of unknown structure that is defined by a conserved sequence and named after the *Trm*2 family that methylates eukaryotic tRNA U54 *a*nd the *Mi*aB family involved in 2-methylthioadenine formation (Anantharaman et al., 2001, 2002). It is also present in several other proteins associated with the translation machinery, and it is predicted to be an RNA binding domain.

The RumA and YbjF subfamilies both contain four conserved cysteine residues in a sequence motif similar to that found in iron-sulfur proteins. Based on the presence of this sequence motif, absorption spectrum, and metal analysis, we had suggested that RumA may carry an iron-sulfur cluster (Agarwalla et al., 2002). Recently, one other RNA modifying enzyme, MiaB, was identified as an iron-sulfur protein (Pierrel et al., 2002). The iron-sulfur cluster is not found in the RUMT subfamily, and its function within a SAM-dependent MTase whose catalytic mechanism does not involve a redox step remains enigmatic.

Here we report the crystal structure of RumA at 1.95 Å. The protein is organized into three domains: an N-terminal TRAM domain, a central domain containing an iron-sulfur cluster, and a C-terminal domain that displays the typical SAM-dependent methyltransferase fold. Comparison of this structure with m^5C MTase li-

*Correspondence: stroud@msg.ucsf.edu

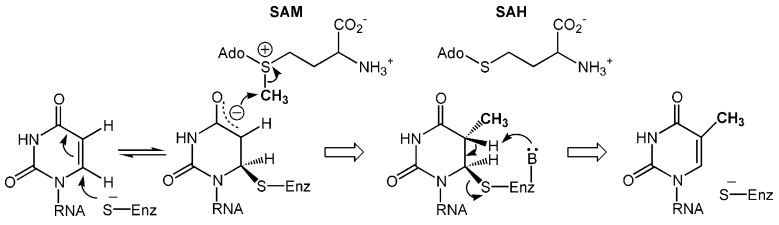

Figure 1. The Catalytic Mechanism of RNA m^{5}U MTases

Adapted from Kealey et al. (1994).

ganded structures allows predictions regarding elements involved in substrate binding and specificity.

Results and Discussion

Overall Structure

The structure of RumA was determined using a single crystal of the selenomethionyl protein (Table 1). The crystal space group is $P2_1$ with one molecule in each asymmetric unit. The final model contains 417 of the 432 residues (Gln15–Arg431). Electron density for the 14 N-terminal residues and the 2 C-terminal residues is not interpreted. These residues are probably disordered and so are excluded from the model. All residues are in the "allowed" region of the Ramachandran plot (Ramakrishnan and Ramachandran, 1965).

The protein is organized into three structural domains (Figure 2). The N-terminal domain (residues 15–74) is the smallest domain and is composed of five β strands. The central domain (residues 75–92 and 125–262) and the catalytic domain (residues 93–124 and 263–431) are juxtaposed, and a concave surface is formed at the interface between them. Figure 2B shows a topology diagram of the RumA structure.

The N-Terminal TRAM Domain Displays an Oligosaccharide/Oligonuleotide Binding Fold

The N-terminal TRAM domain forms a five-stranded antiparallel β barrel (β1↑β2↓β3↑β4↓β5↑) with a Greek key topology. This architecture is characteristic of an oligosaccharide/oligonuleotide binding fold (OB fold). First described in 1993 (Murzin, 1993), the OB fold is found in many superfamilies of DNA/RNA binding domains. A Dali protein fold search revealed that all top hits are of proteins with domains that are known or predicted to bind oligonuleotides. Although the TRAM domain has been predicted to be an RNA binding domain, no structure was previously available. Finding that the RumA TRAM domain has an OB fold supports an RNA binding function for this and other TRAM domains.

All currently available structures of OB fold domains bind to their substrates in a similar manner, though no sequence conservation can be found among them. The main residues involved in binding are located in the

Table 1. Crystallographic Data

Data collection		
Space group	$P2_1$	
Unit cell dimensions (Å)	a = 35.9, b = 99.4, c = 58.5, β = 100.6°	
Data set	**λ1 (Peak)**	**λ2 (Edge)**
Wavelength (Å)	0.9786	0.9788
Resolution (Å)	50–1.95	50–2.2
Observed reflections	632,764	411,102
Unique reflections	29,176	20,711
R_{merge}	9.4% (89.6%)	8.0%(69.3%)
I/σ(I)	18.4(2.6)	21.1(3.0)
Completeness	100%	100%
Phasing		
Phasing power	1.8	1.6
Figure of merit (initial)	0.40	
Figure of merit (after DM)	0.85	
Refinement		
Resolution (Å)	50–1.95	
Reflections	27,660	
Total atoms	3,255	
Water atoms	123	
Overall B factor (Å^2)	41.7	
$R_{working}$	18.9%	
R_{free}	22.3%	
Rmsd bond lengths (Å)	0.019	
Rmsd bond angles (°)	1.66	

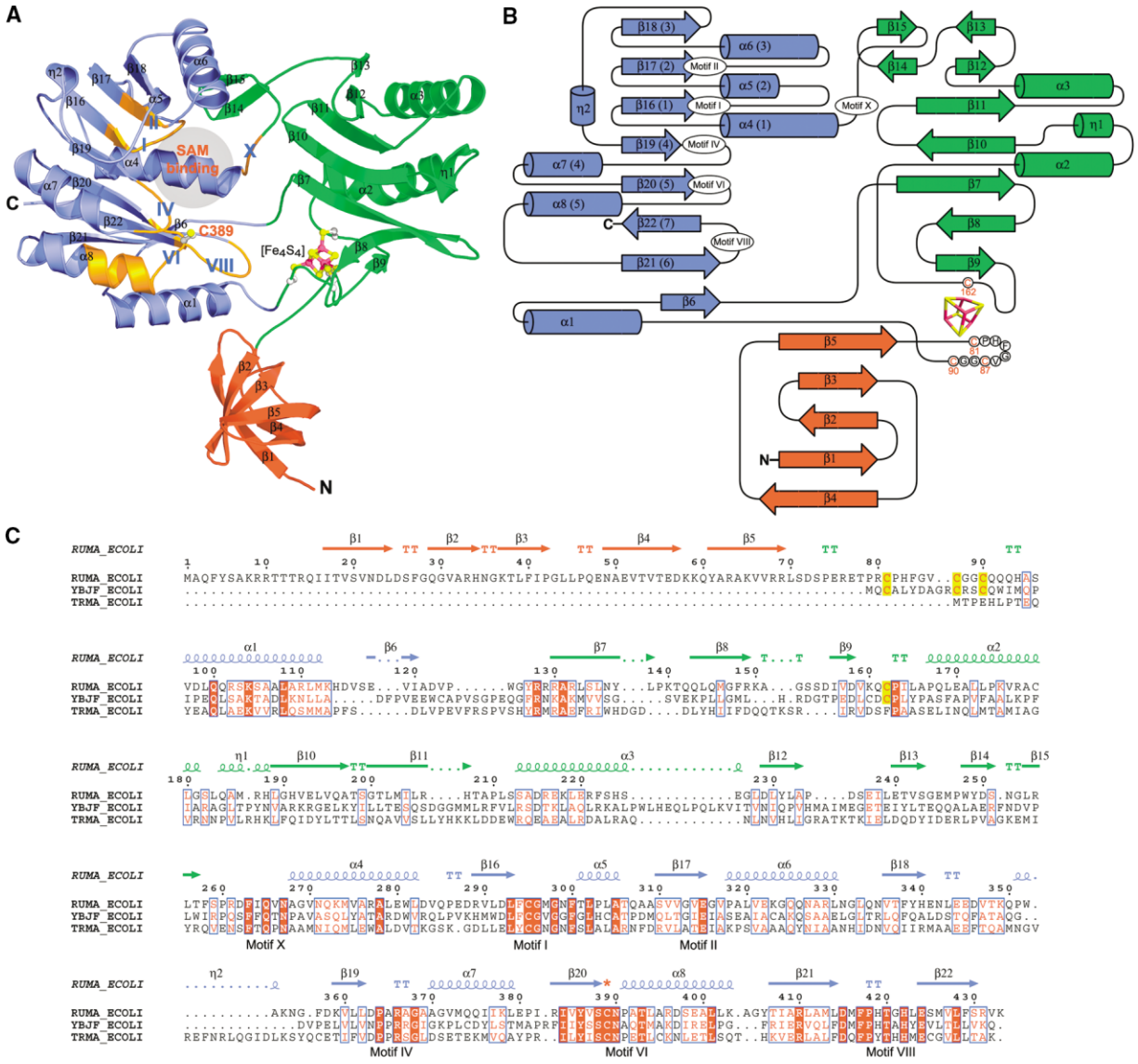

Figure 2. Overall Structure of RumA

The model includes residues 15–431. The N-terminal, central, and C-terminal domains are colored in red, green, and blue, respectively. The α helices and β strands are each numbered sequentially.

(A) Ribbon diagram of RumA. The catalytic cysteine (389), the [Fe_4S_4] cluster, and the side chains of its coordinating cysteines are shown in ball-and-stick model (magenta for Fe, yellow for S, and silver for C). The locations of the conserved MTase motifs are colored in gold and indicated in roman numerals. The potential SAM binding region is also marked.

(B) Planar representation showing the topology of the same model. Classic numbering of the secondary structure elements in the MTase domain is also shown in parentheses.

(C) Sequence alignment of three *E. coli* m^5U MTases. The secondary structure elements of RumA assigned using DSSP (Kabsch and Sander, 1983) are labeled on top of the sequences and correspond with the secondary structure elements shown in (A) and (B). The four [Fe_4S_4]-coordinating cysteines are highlighted in yellow. The conserved motifs and the catalytic cysteine (*) are also marked. The graph was prepared with ESPript (Gouet et al., 1999).

loops connecting β1 and β2, β3 and β4, and β4 and β5 (Murzin, 1993; Suck, 1997). β1, β2, and β3 also provide residues that frequently participate in forming the binding surface. The equivalent residues in the N-terminal domain of RumA are at the front side of the model shown in Figure 2A, indicating a potential RNA binding area.

The N-terminal domain is the least ordered domain in the crystal structure. The average temperature factor of residues in this domain is 69.9 Å^2, whereas the corresponding values for the central and C-terminal domains are 36.1 Å^2 and 37.9 Å^2, respectively. Moreover, when the three domains were treated as separate rigid-body groups in TLS refinement, the resulting anisotropic B factors reveal that the N-terminal domain displays a marked libration displacement, as displayed in Figure 3. The observations of the dissimilar average B factors

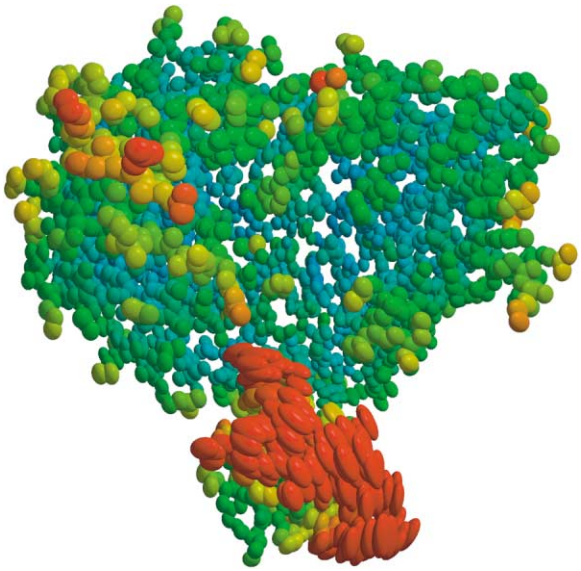

Figure 3. Thermal Ellipsoid Representation of the Anisotropic Temperature Factors

The anisotropic U factors resulting from TLS refinement with the three domains as the TLS groups show marked libration. The ellipsoids were plotted at the 50% probability level and colored according to B_{iso} with a range of 10 (dark blue) to 80 (dark red). The most noticeable motion is the libration of the N-terminal domain. The figure was prepared using RASTEP (Merritt, 1999).

and the anisotropy suggest that the N-terminal domain is flexible in the apo structure. Since this domain appears to be involved in RNA binding, its flexibility is likely to be associated with a conformational change that facilitates binding of the macromolecular substrate, and the shape and orientation of the thermal ellipsoids may indicate the direction of domain motion that accompanies binding to RNA.

The [Fe_4S_4]-Containing Central Domain

RumA contains four conserved cysteine residues (Cys81, Cys87, Cys90, and Cys162) that coordinate an iron-sulfur cluster. In the experimental map, the density of the cluster has the shape of a distorted cube whose geometry matched the shape and dimensions of a [Fe_4S_4] cluster (Liu et al., 2002). The density at each iron atom is continuously connected to one of the four cysteine residues in the experimental map contoured at the 2 σ level. Figure 4 shows the simulated annealing $2F_o - F_c$ omit map for which the cluster was omitted during refinement and Fourier synthesis.

The [Fe_4S_4] binding pocket is formed by an extended loop (75–92) and a β hairpin (147–163). The cluster tethers the two peptide segments through binding of the four cysteines. As a result, the segment containing Cys81, Cys87, and Cys90 is integrated into the rest of the central domain. The interaction between the two segments appears to be critical in maintaining the general scaffold. In order to understand the functional role of the iron-sulfur cluster, we constructed eight single mutants where each of the cluster binding cysteine residues was mutated to either alanine or leucine and a double mutant where cysteine 87 and cysteine 90 were mutated to serine. Although all of these mutant proteins expressed efficiently, they all produced insoluble proteins, precluding any further characterization (data not shown). These observations, however, suggest that the cluster in RumA is required for proper folding of the protein and favor a structural role.

[Fe_4S_4] clusters usually have three accessible redox states: 1+, 2+, and 3+ (Sticht and Rosch, 1998). RumA is silent in electron paramagnetic resonance analysis, indicating most probably a charge of +2 for the cluster and a charge of −2 for the [Fe_4S_4]-Cys_4 center. However, no charged residue lies within 6 Å of the cluster to counterbalance the charge. Rather, the cluster is enclosed mainly by hydrophobic side chains. Besides the coordinating cysteines, seven nonpolar residues and two histidines interact with the cluster. Two of the cluster sulfur atoms form hydrogen bonds with protein atoms (S2 to His83-Nδ1 and S4 to Ile164-N and Gln93-Nϵ2). Two of the chelating cysteine sulfurs form H bonds to main chain nitrogen atoms (Cys81 to His83 and Cys87 to Gly89). Despite these interactions, the cluster is not completely buried. In particular, the surface of S3 of the cluster is partially exposed. It is on the edge of the proposed RNA binding region (see below). Therefore, this prosthetic group is potentially in a position to contact the substrate RNA through hydrogen bonding or van der Waals interactions.

Two glycine residues invariant among the RumA subfamily members appear to be critical in constructing the [Fe_4S_4] binding pocket. Gly147, located in the middle of β8, has a phi angle of 173.6° and a psi angle of 142.0°, well out of the range for a typical β strand residue. As a result, β8 is bent by almost 90° at Gly147. The portion of β8 following Gly147 and β9 forms a hairpin that becomes part of a binding wall that shields the prosthetic group from the solvent (Figure 2A). Therefore, glycine is probably required at position 147 to ensure the proper structural scaffold for the cluster binding residues. Unlike Gly147, the torsion angles of Gly89 (phi = −75.4° and psi = −37.7°) are in a "most favored" region of the Ramachandran plot. Thus, glycine is not required for the conformation of the main chain here. Rather, it is the absence of the Cβ at position 89 that appears to be critical. The Cα of Gly89 is in van der Waals contact with S3, the only solvent-exposed atom in the [Fe_4S_4] cluster, and a Cβ of any other amino acid at this position would block cluster binding and/or interfere with any possible interactions between the RNA and the cluster.

Besides electron transfer, iron-sulfur clusters are also known to have catalytic, structural, regulatory, and sensing functions (Beinert et al., 1997). The absence of an iron-sulfur cluster in the homologous m^5U MTases RUMT and Trm2 makes its participation in the catalysis of the methyl transfer reaction unlikely. Two proteins in which the iron-sulfur cluster does not participate in electron transfer or catalysis may provide insight into other possible roles for this prosthetic group. Like RumA, the cluster in glutamine 5-phosphoribosyl-1-pyrophosphate (PRPP) amidotransferase is surrounded by hydrophobic residues, and a cluster sulfur is exposed to solvent in both cases. The cluster in glutamine PRPP amidotransferase has been proposed to have both a struc-

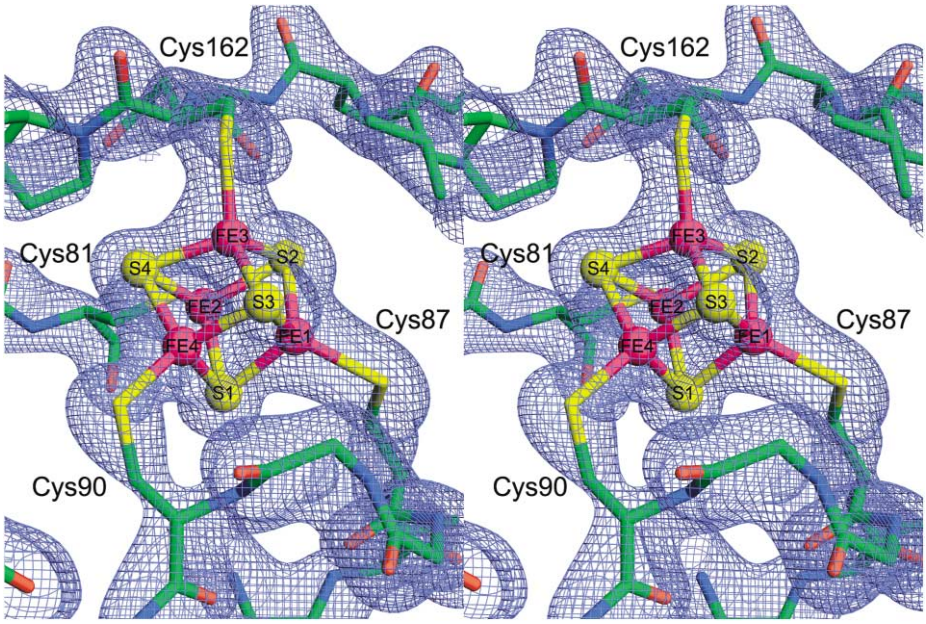

Figure 4. Stereo View of the Sigma-Weighted $2F_o - F_c$ Omit Electron Density around the $[Fe_4S_4]$ Cluster

The map contoured at 1 σ was calculated after the simulated annealing procedure in CNS omitting the coordinates of the cluster. The covalent bonding between the cluster and all four cysteine side chains is revealed by continuous electron density.

tural function and a regulatory role, since oxygen-dependent decomposition of the cluster leads to denaturation of the enzyme (Smith et al., 1994). In endonuclease III, the charge of the $[Fe_4S_4]$-Cys_4 center is counterbalanced by many charged residues, including several conserved positively charged residues. It has been proposed that the cluster has a role in maintaining the fold and presenting these residues to bind DNA phosphate backbone (Fromme and Verdine, 2003; Thayer et al., 1995). The charged environment of the endonuclease III cluster contrasts with the hydrophobic environment of the RumA cluster, and the positively charged residues in RumA that presumably bind the nucleic acid backbone are more distant from the cluster than those in endonuclease III.

Besides the iron-sulfur cluster, the most prominent structural feature of the central domain is a six-stranded β sheet (β8$\downarrow$ β7$\uparrow$ β10$\downarrow$ β11$\uparrow$ β12$\uparrow$ β13$\downarrow$) combined with three helices. A DALI search failed to detect any known fold with significant similarity to this 103-residue-long fold. Sequence alignment of proteins in the RumA subfamily shows that this segment is the least conserved region of the whole protein. It contains residues that form part of a positively charged surface, indicating it may be involved in RNA binding (see below).

The Catalytic Domain Shows a Typical SAM-Dependent MTase Fold with an Extension

The C-terminal domain is the methyl transferring domain and displays a mixed eight-stranded β sheet which is flanked on both sides by a total of six α helices and one 3_{10} helix. The first seven strands (β16–β22) exhibit the characteristic topology of the SAM-dependent MTase fold, where the first six strands of the sheet (β16–β21) are parallel and each is preceded by an antiparallel helix. The only antiparallel strand, strand 7 (β22), follows right after strand 6 (β21) and loops back to insert between strand 6 and strand 5 (β20).

In RumA, the core SAM binding fold is extended by the segment defined by residues 93–124, containing α1 and β6. This segment crosses over from the central domain and becomes integrated into the catalytic domain, where β6 is the eighth β strand in the sheet. This insert apparently has little to do with SAM binding or catalysis. However, it might function to improve the stability of the MTase fold as well as the overall stability of the three-domain structure, because the central and the C-terminal domains are now linked through three covalent linkages, two more than they would have without this crossover insert. The crossover increases the connection between the central domain and the catalytic domain, tethering these two domains together more tightly, which in turn might facilitate the communication between the two domains and the delivery of the substrate to the active site upon RNA binding.

There are up to ten conserved motifs (I–X) in SAM-MTases as revealed by previous structural and sequence analyses (Malone et al., 1995; Posfai et al., 1989; Reid et al., 1999; Schluckebier et al., 1995). A sequence alignment using RumA homologs from prokaryotes, eukaryotes, and archea identified six conserved blocks corresponding to motifs X, I, II, IV, VI, and VIII of m^5C MTases (Figures 2 and 5). All known SAM-MTases share several common features in binding SAM (Fauman et al., 1999). These features are also found in RumA. They include the glycine-rich loop (Gly296 and Gly298) in motif I and a negatively charged residue (Glu315) in motif II (Figure 5). Thus, the SAM binding environment in RumA appears to be the same as that in most other SAM-MTases.

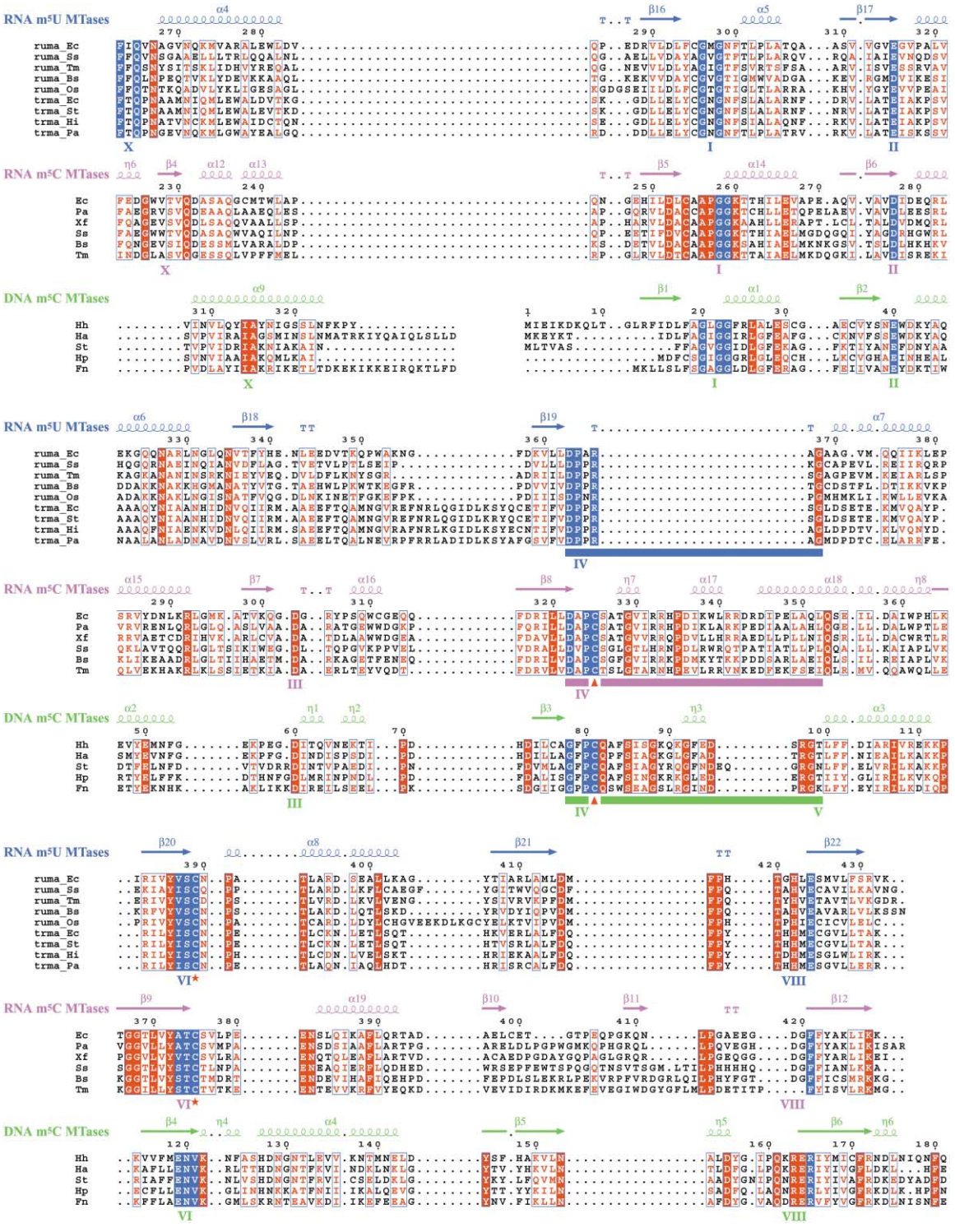

Figure 5. Sequence Alignment of the Catalytic Domains of Three 5-Methylpyrimidine MTase Families

Sequence alignment within each family was performed by CLUSTALW. The secondary structural elements that are colored and labeled on top of each family are those of RumA (RNA m^{5}U MTases, blue), Fmu (RNA m^{5}C MTases, magenta), and *M. Hha*I (DNA m^{5}C MTases, green). The conserved motifs are marked below the respective sequences with the same color as the secondary structural elements. Alignment between families is adjusted to reflect the 3-D structural conservation. The plot was generated by ESPript using the default color setting (white letters highlighted in red for complete conservation, and red letters for residues with similarity scores greater than 0.7), except that the conserved residues discussed in the text are colored blue. The conserved cysteines in motifs IV (m^{5}C MTases) and VI (RNA MTases) are indicated by red triangles and red stars, respectively. The segment marked with a rectangular bar shows large differences among the three families. RumA, Fmu, and *M. Hha*I contain 6, 31, and 22 residues in this segment. (The segment starts with motif IV and ends at the first

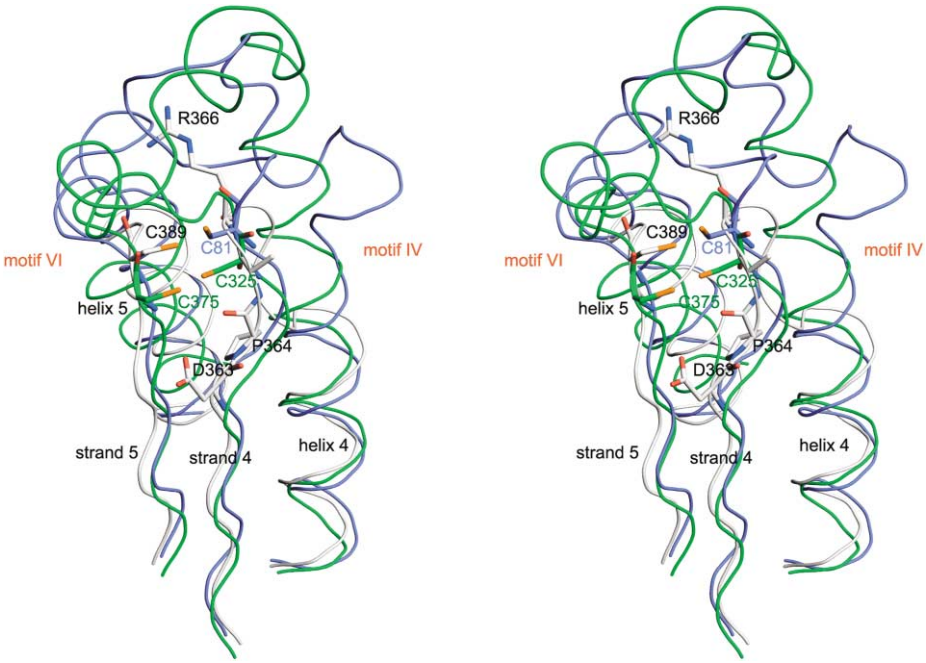

Figure 6. Stereo View of Motifs IV and VI of Superposed RumA, Fmu, and *M. Hha*I

The structures were aligned using the LSQKAB program in the CCP4 package (CCP4, 1994; Kabsch, 1976) using the Cα atoms of equivalent residues, as identified by the DALI server. Motifs IV and VI and their flanking strands and helices are shown in ribbons. The ribbon and the carbon atoms are colored in silver (RumA), green (Fmu), and blue (*M. Hha*I). The rest of the atoms are colored by atom type (red for O, blue for N, and orange for S). The catalytic cysteines in RumA (Cys389) and Fmu (Cys375) are from motif VI, whereas the catalytic cysteine in *M. Hha*I (Cys81) is from motif IV. They are in close proximity despite the differences of the nucleotide substrates. The segments marked by rectangular bars in Figure 5 show striking divergence in length and orientation. The second conserved cysteine in Fmu (Cys325) is also shown. The side chains of motif IV following strand 4 in RumA (DPAR) are also shown.

Adaptation between DNA and RNA MTase Catalytic Sites

To uncover the common features shared by 5-methylpyrimidine MTases and elements that might dictate the basis of different substrate specificity, we compared the catalytic domains of RumA and the DNA/RNA m^5C MTases. Structures of m^5C MTases have been determined for three enzymes: DNA m^5C MTase *M. Hha*I (O'Gara et al., 1996), DNA m^5C MTase *M. Hae*III (Reinisch et al., 1995), and RNA m^5C MTase Fmu (Foster et al., 2003).

Both m^5U and m^5C MTases feature a catalytic cysteine that carries out nucleophilic attack at C-6 of the pyrimidine ring and activates C-5 for methylation. In the DNA m^5C MTases, this catalytic cysteine is in motif IV (Cys81 in *M. Hha*I and Cys71 in *M. Hae*III) and is the fourth residue following strand 4 of the core β sheet (Figure 5, red triangle) (O'Gara et al., 1996). This motif IV cysteine is conserved across all subfamilies of DNA m^5C MTases (Posfai et al., 1989). Curiously, this motif IV cysteine is also conserved in the RNA m^5C MTases (Reid et al., 1999). However, a variety of biochemical and mutagenic analyses have shown that a conserved cysteine in motif VI, but not the one in motif IV, is the catalytic nucleophile for both the RNA m^5C and m^5U MTases (Figure 5, red star) (Kealey and Santi, 1991; Liu and Santi, 2000). In RumA, the motif VI cysteine is Cys389, and it is the first residue following strand 5. Figure 6 shows the close proximity of these different cysteines after superposition of three MTases (RumA, Fmu, and *M. Hha*I). The distance between Cys389-Sγ of RumA and Cys81-Sγ of *M. Hha*I is only 1.94 Å, whereas the distance between Cys375-Sγ of Fmu and Cys81-Sγ of *M. Hha*I is 3.41 Å. Thus, despite the large distances in sequence between the motif IV cysteine and motif VI cysteines, they are spatially very close in the 3-D folds of the DNA and RNA 5-methylpyrimidine MTases. The structure of *M. Hha*I used for this comparison is that of a complex with the products, *S*-adenosylhomocysteine and methylated DNA, whereas the Fmu structure is a complex with just the cofactor SAM. Considering the conformational change that may accompany substrate binding, relatively minor adjustments would be required to position C-6 of the substrate appropriately relative to the differ-

structurally alignable residue.) Sequences of RNA m^5U MTases include the ruma and trma (RUMT) subfamilies, which are specified in their names. The two-letter species abbreviations are: Ec, *Escherichia coli*; Ss, *Synechocystis sp.*; Tm, *Thermotoga maritima*; Bs, *Bacillus subtilis*; Os, *Oryza sativa*; St (in RNA m^5U MTases), *Salmonella typhimurium*; Hi, *Haemophilus influenzae*; Pa, *Pseudomonas aeruginosa*; Xf, *Xylella fastidiosa*; Hh, *Haemophilus haemolyticus*; Ha, *Haemophilus aegyptius*; St (in DNA m^5C MTases), *Streptococcus thermophilus*; Hp, *Helicobacter pylori*; Fn, *Fusobacterium nucleatum*.

ent catalytic cysteines. Thus, the DNA and RNA 5-methylpyrimidine MTases have adapted to employ different motifs to perform the same catalytic function.

The Structural Basis for Selectivity in m^5U and m^5C MTases

There are profound structural variations in motifs IV, VI, VIII, and X between the DNA/RNA MTases that most probably encode selectivity for the three-dimensionally folded target nucleic acid sequence. Motif IV, following strand 4 (β19 in RumA), contains a four to six amino acid signature that distinguishes each of the families of pyrimidine MTases involved in RNA metabolism (Anantharaman et al., 2002). Motif IV of each of these classes of MTases contains a conserved proline, but its position is not conserved between the RNA MTases and the DNA MTases, and the adjacent residues are not conserved between the m^5U and m^5C MTases. In the RNA MTases, the conserved proline is the second residue of the motif, and in the m^5U MTases, it is preceded by a conserved aspartate (Asp363 in RumA). This aspartate may be important in positioning the uracil or in some other aspect of catalysis. Proline places a torsional constraint on the main chain and may be important in correctly positioning the neighboring residues that are critical in recognition or catalysis. Besides the differing catalytic function of motif IV, the segment marked by the rectangular bars in Figure 5 represents the largest structural divergence in the catalytic domain among the available 5-methylpyrimidine MTase structures (Figures 5 and 6).

The position of the DNA MTase catalytic cysteine is occupied by a conserved arginine in the m^5U MTases (Arg366 in RumA) (Figure 5). Since this arginine is not a nucleophile, it is likely to have some other function. Arginine residues are often employed in binding to the nucleic acid backbone, and we speculate that this arginine may have a role such as orienting the substrate.

Motif VI of DNA m^5C MTases is involved in target base recognition. It contains a conserved glutamate (Glu119 in *M. Hha*I) that hydrogen bonds with N4 of the target cytosine (O'Gara et al., 1996). On the contrary, this motif in RNA MTases harbors the catalytic nucleophile (Cys389 in RumA). Therefore, the role of motif VI is changed from catalysis in RNA MTases to substrate binding in DNA MTases.

Motifs VIII and X also display large structural divergences among the 5-methylpyrimidine MTase families. In *M. Hha*I, an invariant arginine (Arg165) in motif VIII interacts with the base and the phosphate moiety of the target cytosine (O'Gara et al., 1996). RumA and Fmu have different residues (Glu424 in RumA and Phe421 in Fmu) at this position that are also invariant in their own families (Figure 5). The differences in their chemical properties suggest that these residues have distinctive roles in protein functions. Motif X in RumA contains two residues, Phe263 and Gln265, in the loop preceding helix 1 (α4) that are invariant in every m^5U MTase subfamily but are not present in m^5C MTases (Figure 5). Therefore, these residues might have important roles unique to the RNA m^5U MTases.

The pattern of high conservation within subfamilies, and the distinctive variations between subfamilies suggest that motifs IV, VI, VIII, and X have key roles in specificity for folded tertiary structures of substrate RNAs versus duplex DNA, as well as for targeting uracil versus targeting cytosine. In the complex of *M. Hha*I and DNA, the segment containing motif IV (Figure 5, rectangular bars) is inserted into the minor groove of the dsDNA substrate. In RNA MTases, this segment is oriented differently, indicating possible dissimilar conformations of the RNA substrates. The residence of the catalytic cysteine in motif IV versus motif VI is characteristic of the DNA MTases versus RNA MTases. This must be accompanied by adaptations in the positions of residues that correctly position the base within the binding pocket and discriminate between the two substrate bases, uracil and cytosine, that have different hydrogen bonding properties. The structural and sequence comparison of these three subfamilies of 5-methylpyrimidine MTases reveals several residues that are unique to m^5U MTases and are potential candidates for base discrimination.

RNA Binding Surface

The structure determined here is of the apo protein. Nevertheless, there is good evidence for the location of the RNA binding site. We propose that the RNA binds to the surface of the N-terminal and central domains of the protein close to the junction of the three domains. This hypothesis is based on the following evidence.

The electrostatic potential surface of RumA (Figure 7) shows that one side is highly charged and has a distinctive profile, whereas the opposite side is relatively featureless. From the front side as viewed in Figures 2 and 7, the C-terminal catalytic domain is dominated by negatively charged residues, which are concentrated in area C. This is consistent with its role in binding the positively charged SAM cofactor for catalysis. In contrast, the region on the central domain that forms the interdomain cleft (area B) along with the N-terminal domain on the same side (area A) is covered with predominately positively charged residues. Area A coincides with the proposed RNA binding site of the OB fold. Areas A and B form the largest continuous, positively charged surface on the protein and lie in the groove between the central and catalytic domains. A stretch of conserved positively charged residues, including Arg128 and Arg130, is located on the surface in this region. The presence of these conserved residues and the electrostatic properties of this region are consistent with its role in binding the negatively charged RNA substrate.

Comparison of 5-methylpyrimidine MTase strucutres supports the proposed RNA binding site. In the structures of the DNA m^5C MTases *M. Hha*I and *M. Hae*III, the locations of the DNA binding domains relative to the catalytic domains are similar to that of the central domain of RumA. The predicted RNA binding domain in Fmu is also found at this location. These observations support the role of the central domain as an RNA binding domain.

In conclusion, RumA is the first three-dimensional structure for any RNA-modifying protein containing a $[Fe_4S_4]$ iron-sulfur cluster. Possible roles for this prosthetic group in RumA have been suggested: (1) It might

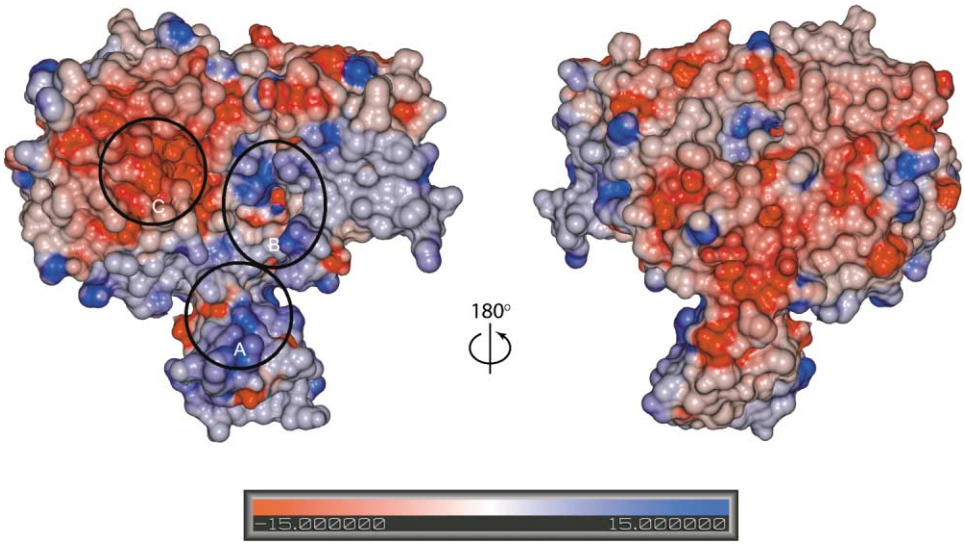

Figure 7. Electrostatic Potential Surface of RumA

The molecular surface was calculated with MSMS (Sanner et al., 1996) using a probe radius of 1.35 Å. An electrostatic potential map was computed with GRASP (Nicholls et al., 1991) using the Poisson-Boltzmann method. The cluster was assigned +2 charge, and the four coordinating cysteines were each assigned −1 charge. The surface of the front side (left) shows three highly charged areas (enclosed and labeled A, B, and C), whereas the backside (right) lacks any continuous stretch of positively charged surface. The color ramp and the scale are shown at the bottom, with blue for positive (+15 k_BT), white for neutral (0 k_BT), and red for negative (−15 k_BT), where k_B is the Boltzmann's constant and T is the temperature.

have a structural role, stabilizing the protein by holding together two structural elements. (2) The cluster is proximal to the proposed RNA binding site and is partially exposed to solvent, and thus might participate in RNA binding. (3) The cluster might regulate the stability or function of the protein in response to the changes in cellular environment (Rouault and Klausner, 1996). A totally independent function of the iron-sulfur cluster cannot be ruled out, and its function in RumA remains unclear.

TRAM is a well known sequence motif, but its structure had not previously been determined. Here we demonstrate that a TRAM sequence adopts a common structural fold, the OB fold, which is known to bind nucleic acids. The catalytic domain of RumA displays a canonical methyltransferase fold with an additional β strand which crosses over from the central domain, integrating these two domains. The catalytic Cys389 is located in a different motif than that in the DNA m^5C MTases. The shape, the surface electrostatic potential of the protein, and the location of the conserved positively charged residues allow us to propose the RNA binding region of the protein.

RumA catalyzes the methylation of U1939 of the 23S rRNA with high specificity. It does not catalyze the methylation of U1939C mutant 23S rRNA (Agarwalla et al., 2002). Therefore, the enzyme not only recognizes the specifically folded RNA structure, presumably through the RNA binding domains, but it also differentiates between the two pyrimidine nucleotides, uridine and cytidine, at the site of the reaction. Structural determination of the protein-RNA complex will provide insight into the detailed mechanism of catalysis and substrate recognition.

Experimental Procedures

Protein Crystallization and Data Collection

E. coli RumA was cloned, expressed, and purified as described (Agarwalla et al., 2002). The purified protein was dialyzed against 10 mM potassium phosphate containing 100 mM NaCl and 10% glycerol (pH 7.0). Crystals were grown using the hanging drop vapor diffusion method. Protein (3 mg/ml) was mixed with an equal volume of well buffer (50 mM Tris-Cl [pH 9.0–9.5], 10 mM $NiCl_2$, 12.5% glycerol, and 20% polyethylene glycol 2000 monomethyl ether). The reservoir contained 500 μl well buffer and was covered with 200 μl silicon oil before sealing. The plate-shaped crystals grew to the optimal size (80 × 40 × 10 μm^3 on average) in 2–3 days. A distinct dark yellow color of RumA was obvious in the thicker plates. To prevent the crystals from dissolving and the protein from precipitating, the crystals were frozen in liquid nitrogen after 3 days of growth.

Diffraction data were collected on frozen crystals at the Advance Light Source (ALS, Lawrence Berkeley Lab, CA). Two wavelengths (peak and inflection) were set based on a fluorescence scan. For each wavelength, 360 one-degree oscillation frames were collected, and the crystal was rotated 180° after every 30 frames. The crystals are space group $P2_1$ with a = 35.9 Å, b = 99.4 Å, c = 58.5 Å, and β = 100.6°. Data were processed and scaled using DENZO and SCALEPACK (Otwinowski and Minor, 1997).

Structure Determination and Refinement

The phases were determined by single-wavelength anomalous dispersion methods using the peak wavelength data set collected for the selenomethionyl protein. The positions of all eleven selenium atoms were determined. Seven sites were identified from the Patterson map, and the remaining four sites were then located in double-difference maps. Phase refinement and density modification were carried out with SHARP (de La Fortelle and Bricogne, 1997). Approximately 70% of the whole molecule was built into a map calculated with the modified phases with assistance of ARP/wARP (Perrakis et al., 1999). Remaining residues were built manually using Quanta (Accelrys, San Diego, CA). Structural refinement was carried out using CNS (Brunger et al., 1998) and REFMAC5 (CCP4, 1994; Murshudov et al., 1997) with refinement of TLS parameters (Winn

et al., 2001). The stereochemical quality of the final structure was verified by PROCHECK (Laskowski et al., 1993). Figures were prepared with PyMOL (DeLano Scientific) unless noted.

Acknowledgments

We thank Janet Finer-Moore for help with structure determination and Pat Greene for critical reading of the manuscript. We also thank Luke Rice for help on CNS. This work was supported by United States Public Health Service, National Institutes of Health Grant GM51232 (to R.M.S.).

Received: August 14, 2003
Revised: October 31, 2003
Accepted: November 13, 2003
Published: March 9, 2004

References

Agarwalla, S., Kealey, J.T., Santi, D.V., and Stroud, R.M. (2002). Characterization of the 23 S ribosomal RNA m^{5}U1939 methyltransferase from *Escherichia coli*. J. Biol. Chem. *277*, 8835–8840.

Anantharaman, V., Koonin, E.V., and Aravind, L. (2001). TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. FEMS Microbiol. Lett. *197*, 215–221.

Anantharaman, V., Koonin, E.V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res. *30*, 1427–1464.

Beinert, H., Holm, R.H., and Munck, E. (1997). Iron-sulfur clusters: nature's modular, multipurpose structures. Science *277*, 653–659.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N.S., et al. (1998). Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In Methods in Enzymology, J.C.W. Carter and R. M. Sweet, eds. (New York: Academic Press), pp. 472–494.

Decatur, W.A., and Fournier, M.J. (2002). rRNA modifications and ribosome function. Trends Biochem. Sci. *27*, 344–351.

Del Campo, M., Kaya, Y., and Ofengand, J. (2001). Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. RNA *7*, 1603–1615.

Fauman, E.B., Blumenthal, R.M., and Cheng, X. (1999). Structure and evolution of AdoMet-dependent methyltransferases. In S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions, X. Cheng and R.M. Blumenthal, eds. (Singapore: World Scientific Publishing), pp. 1–38.

Foster, P.G., Nunes, C.R., Greene, P., Moustakas, D., and Stroud, R.M. (2003). The first structure of an RNA m^{5}C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate. Structure (Camb) *11*, 1609–1620.

Fromme, J.C., and Verdine, G.L. (2003). Structure of a trapped endonuclease III-DNA covalent intermediate. EMBO J. *22*, 3461–3471.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.

Gustafsson, C., Reid, R., Greene, P.J., and Santi, D.V. (1996). Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. Nucleic Acids Res. *24*, 3756–3762.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. Acta Crystallogr. A *32*, 922–923.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

Kealey, J.T., and Santi, D.V. (1991). Identification of the catalytic nucleophile of tRNA (m^{5}U54)methyltransferase. Biochemistry *30*, 9724–9728.

Kealey, J.T., Gu, X., and Santi, D.V. (1994). Enzymatic mechanism of tRNA (m^{5}U54)methyltransferase. Biochimie *76*, 1133–1142.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Limbach, P.A., Crain, P.F., and McCloskey, J.A. (1994). Summary: the modified nucleosides of RNA. Nucleic Acids Res. *22*, 2183–2196.

Liu, Y., and Santi, D.V. (2000). m^{5}C RNA and m^{5}C DNA methyl transferases use different cysteine residues as catalysts. Proc. Natl. Acad. Sci. USA *97*, 8263–8265.

Liu, L., Nogi, T., Kobayashi, M., Nozawa, T., and Miki, K. (2002). Ultrahigh-resolution structure of high-potential iron-sulfur protein from Thermochromatium tepidum. Acta Crystallogr. D Biol. Crystallogr. *58*, 1085–1091.

Malone, T., Blumenthal, R.M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J. Mol. Biol. *253*, 618–632.

Merritt, E.A. (1999). Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Crystallogr. D Biol. Crystallogr. *55*, 1109–1117.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J. *12*, 861–867.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

O'Gara, M., Klimasauskas, S., Roberts, R.J., and Cheng, X. (1996). Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for *Hha*I methyltransferase-DNA-AdoHcy complexes. J. Mol. Biol. *261*, 634–645.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, J.C.W. Carter and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Pierrel, F., Bjork, G.R., Fontecave, M., and Atta, M. (2002). Enzymatic modification of tRNAs: MiaB is an iron-sulfur protein. J. Biol. Chem. *277*, 13367–13370.

Posfai, J., Bhagwat, A.S., Posfai, G., and Roberts, R.J. (1989). Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res. *17*, 2421–2435.

Ramakrishnan, C., and Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. Biophys. J. *5*, 909–933.

Reid, R., Greene, P.J., and Santi, D.V. (1999). Exposition of a family of RNA m^{5}C methyltransferases from searching genomic and proteomic sequences. Nucleic Acids Res. *27*, 3138–3145.

Reinisch, K.M., Chen, L., Verdine, G.L., and Lipscomb, W.N. (1995). The crystal structure of *Hae*III methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. Cell *82*, 143–153.

Rouault, T.A., and Klausner, R.D. (1996). Iron-sulfur clusters as biosensors of oxidants and iron. Trends Biochem. Sci. *21*, 174–177.

Rozenski, J., Crain, P.F., and McCloskey, J.A. (1999). The RNA Modification Database: 1999 update. Nucleic Acids Res. *27*, 196–197.

Sanner, M.F., Olson, A.J., and Spehner, J.C. (1996). Reduced sur-

face: an efficient way to compute molecular surfaces. Biopolymers *38*, 305–320.

Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. J. Mol. Biol. *247*, 16–20.

Smith, J.L., Zaluzec, E.J., Wery, J.P., Niu, L., Switzer, R.L., Zalkin, H., and Satow, Y. (1994). Structure of the allosteric regulatory enzyme of purine biosynthesis. Science *264*, 1427–1433.

Sticht, H., and Rosch, P. (1998). The structure of iron-sulfur proteins. Prog. Biophys. Mol. Biol. *70*, 95–136.

Suck, D. (1997). Common fold, common function, common origin? Nat. Struct. Biol. *4*, 161–165.

Thayer, M.M., Ahern, H., Xing, D., Cunningham, R.P., and Tainer, J.A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. EMBO J. *14*, 4108–4120.

Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. *57*, 122–133.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. Science *292*, 883–896.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank under accession code 1UWV.